A FUZZY APPROACH TO GROUPING BY POLICYHOLDER AGE IN GENERAL INSURANCE

by

R J VERRALL and Y H YAKOUBOV

Actuarial Research Paper No. 104

Department of Actuarial Science and Statistics
City University
London

January 1998

ISBN 1 901615 22 7

A Fuzzy Approach to Grouping by Policyholder Age in General Insurance*

R.J.Verrall[1] and Y.H.Yakoubov

Abstract
This paper considers the treatment of policyholder age in general insurance. In practice, this is often treated as a factor with the number of levels requiring that the individual ages of the policyholders are grouped. Although the groups are usually defined by the existing underwriting structure, it should be investigated as part of any premium rating exercise that uses a model to assess past claims experience. It is possible that an incorrect grouping by policyholder age could bias the results of the risk premium estimation. On the other hand it may not be computationally feasible to use separate ages in the premium model, making some form of grouping necessary. In this paper, we specify a data-based procedure for grouping by age, using fuzzy set theory. An example is given which illustrates how the method can be used in practice.

Keywords: Fuzzy Set Theory; General Insurance; Premium Rating.

*The authors gratefully acknowledge the financial support provided by the City University Actuarial Research Club without which this research would not have been completed.

---

[1] Department of Actuarial Science and Statistics
City University
Northampton Square
London
EC1V 0HB

1. Introduction

Policyholder age is one of the important underwriting factors used in many general insurance policies, such as motor insurance. Data are usually available subdivided according to (for example) age last birthday, but policyholders are often rated according to which of a set of age groups that they fall in. For example, Brockman and Wright (1991) use 8 levels:

17-18, 19-21, 22-24, 25-29, 30-34, 35-44, 45-54, 55+.

It is usual practice for the boundaries and sizes of the age groups to be fixed *a priori*, usually based on the underwriter's or actuary's judgement, or simply by sticking with the previous structure. However, these assumptions represent part of the specification of the model used to investigate the claims experience and to set the relative premium levels. An investigation into the appropriateness of the age groups should therefore form part of any investigation of a general insurance portfolio. It is possible that an inappropriate choice for the age factors could bias the results from the premium rating model, leading to misleading estimation of risk based on historical data.

This paper shows how fuzzy set methods can be used to give an indication of suitable age groupings based on past data. The advantage of this method is that the results are data-dependent, rather than being entirely subjective. It is intrinsic to the fuzzy approach that there is still an element of subjectivity, but we believe that the results of the fuzzy methods can lead to greater confidence in the groupings. It may be that a decision is taken to retain the historical age groupings, but the method in this paper allows the suitability of these groupings to be assessed.

There are other approaches that could be taken to the treatment of policyholder age. For example, parametric or non-parametric smoothing methods, as employed in the graduation of life tables could be used: see Renshaw (1991) and Verrall (1996) for descriptions of the applications of these models to life tables, within the framework of generalised linear models. A useful overview of the applications of generalised linear models in actuarial science is given by Haberman and Renshaw (1996). However, this paper will consider rating based on the grouping of ages, since this is very often the approach taken in practice. There are a number of reasons for retaining the grouping procedure: it can lead to simple rating structures, it is the familiar method, and it alleviates problems with sparse data. The latter problem can be overcome if, for example, a simple parametric model is used, but this may not be appropriate.

The approach taken here uses fuzzy set theory, which was introduced by Zadeh (1965). In the actuarial literature, the theory was first used by DeWitt (1982) and Lemaire (1990). Ostaszewski (1993) has investigated the possible applications of fuzzy methods in actuarial science, and more recently, several other papers have appeared (e.g. Cummins and Derrig, 1993,1997, Derrig and Ostaszewski, 1994, 1995 and Young, 1995). These papers have covered a wide range of contexts including health underwriting, pricing of general insurance business, asset allocation and marketing. A more recent comprehensive review of fuzzy techniques with actuarial applications is given by Yakoubov and Haberman (1998).

We are not aware of other attempts having been made to use statistical or other methods as aids to the determination of suitable age groups. Our aim is to group individual ages into "clusters" which will form the rating groups for policyholder age. It is clearly necessary to restrict such groupings to adjacent ages, and this requirement will limit the amount of clustering which is possible. Fuzzy clustering is very suitable for this purpose, and can be implemented using the "fuzzy c-means clustering algorithm". This algorithm was first applied in the actuarial literature by Derrig and Ostaszewski (1994) and it should be noted that other methods of fuzzy clustering exist. The algorithm is described in the main part of this paper, and it is then applied to a set of general insurance data.

The paper is set out as follows. Section 2 contains a brief introduction to fuzzy set theory, emphasising the aspects which are relevant to this paper. Section 3 sets out the fuzzy c-means clustering algorithm. Section 4 contains an illustration and section 5 contains the conclusions.

2. Fuzzy Set Theory

We now give a brief introduction to the ideas from fuzzy set theory which will be needed in the following sections. It should be noted that in some cases fuzzy set theory is not used for inference, but rather in order to reach a decision on the basis of fuzzy information. An example of this type of application is Horgby *et al* (1997) where a fuzzy expert system is defined in order to reach an underwriting decision in the case of people with diabetes. In this paper, we use fuzzy set theory to make inferences from the data, where the resulting inference is fuzzy. We do not give a full exposition of the theory of fuzzy sets. For a fairly complete introduction, the reader is referred to Zimmerman (1991).

In the application considered in this paper, the aim is to decide to which group each individual age should be allocated. Considering the event that a particular age belongs to a particular group, it can be seen that the evidence from the data about this event is unlikely to be conclusive. In other words, we will be uncertain about whether this event has occurred or not, and this uncertainty can be quantified by using the concept of fuzziness. Thus, if the individual age under consideration is x, and the age group is denoted by A, we can quantify the degree of membership of x in A by $\mu_A(x)$. This requires the generalisation of the definition of a set (a "crisp" set) to a fuzzy set, as follows. Given a collection of objects and the universe of discourse, U, a fuzzy set A is defined by

$$\mu_A : U \to M$$

where μ_A is the membership function of A
and M is an ordered set.
Note that M is usually defined as the unit interval [0,1]. This will be used throughout this paper. Thus, for a crisp set A, the membership function is defined as

$$\mu_A(x) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A \end{cases}$$

The two extreme values, 0 and 1, represent the lowest and highest degrees of membership, respectively. The degree of membership can be interpreted as the truth value of the statement "x is a member of A". In the application to grouping by policyholder age, the membership function will indicate to which group(s) each of the individual ages should be considered as belonging (noting that this may not be a conclusive decision). In order to interpret the outcome of the algorithm which gives estimates of the membership function for each age, it is useful to discard possibilities where the membership function is very low. This can be achieved by using the notion of an α-cut of a fuzzy set, A, which is defined as the (crisp) set A_α, where

$$A_\alpha = \{x \in U : \mu_A(x) \geq \alpha\}.$$

This set contains the elements for which there is (at least) an α% belief that they are in A.

Considering the application to grouping by policyholder age, the problem is to choose a number of groups into which the policyholder age will be partitioned and then to assess to which group each age belongs. For ease of notation, denote the individual ages by $\{i : i = 1,2,\ldots,n\}$ and the groups by $\{k : k = 1,2,\ldots,c\}$, where we would expect $c << n$. The assessment of appropriate age groupings requires estimates of the elements of the matrix M, defined by

$$\mathrm{M} = \{\mu_{ki} : k = 1,2,\ldots,c; i = 1,2,\ldots,n\}$$

where μ_{ki} is the degree of membership of the ith individual age in the kth group. This is a "fuzzy c-partition" (Bezdek, 1981) if it satisfies:

1). $\mu_{ki} \in [0,1], \quad k = 1,2,\ldots,c; i = 1,2,\ldots,n$

2). $\sum_{k=1}^{c} \mu_{ki} = 1, \quad i = 1,2,\ldots,n$

3). $0 < \sum_{i=1}^{n} \mu_{ki} < n, \quad k = 1,2,\ldots,c.$

Elements can belong to two or more clusters to some extent, determined by the membership functions. Note that the boundaries of the groups are not determined *a priori*. Note also that we would expect adjacent ages to lie in the same, or adjacent groups. This is not specified as part of the model, but will be considered after the matrix M has been estimated and an appropriate α-cut has been applied. If the results do not indicate that a relatively smooth transition between the groups is possible, then it is clear that age grouping may not be appropriate (or at least that the groups cannot be determined from the data). The next section shows how the values of the membership functions can be estimated from the data.

3. The Fuzzy c-means Algorithm

In order to calculate values of the membership functions for each individual age, an objective function is required to which an optimisation criterion can be applied. In general, we may have a number of different features of the data which should be used to determine the age groups: for example, different categories of claims, claims frequency and claims severity, etc. In order to be as general as possible, we define the data set by

$$X = \{x_i : i = 1,2,\ldots,n\}$$

where $x_i \in \Re^p$ and each observation consists of p features:

$$x_i = \{x_{ij} : j = 1,2,\ldots,p\}.$$

Clearly, if we consider two ages, l and m (say), the decision on grouping will be based on a measure of their dissimilarity, which is taken to be their distance apart

$$d(x_l, x_m) = \|x_l - x_m\|$$

for a suitably defined norm on X, $\|.\|$. The degrees of membership of each individual age in each age group are defined by minimising $z_m(\mathrm{M}, \upsilon)$ over M and υ, where

$$z_m(\mathrm{M}, \upsilon) = \sum_{k=1}^{c} \sum_{i=1}^{n} (\mu_{ki})^m \|x_i - \upsilon_k\|^2 \qquad (3.1)$$
$$\upsilon = \{\upsilon_k : k = 1,2,\ldots,c\}$$

and υ_k is the "centre" of the kth cluster, to be estimated in the optimisation procedure (see below). The exponential weight m (m>1) reduces the influence of noise in the membership values in relation to the clustering criterion. The larger m, the more weight is assigned to elements with a higher degree of membership, and the less to those with a lower degree of membership. As $m \to \infty$, the membership function tends towards the constant value of $\frac{1}{c}$, indicating that each element is assigned to each cluster with the same degree of membership. It is obviously preferable to have M less fuzzy, and, usually, m is taken to be 2. To motivate the form of $z_m(\mathrm{M}, \upsilon)$, note that each term increases as μ_{ki} increases, and as $\|x_i - \upsilon_k\|^2$ increases. Thus, minimising $z_m(\mathrm{M}, \upsilon)$ will assign low membership values when $\|x_i - \upsilon_k\|^2$ is large, and *vice versa.*

It can be shown (see Bezdek, 1981) that a local minimum of (3.1) is obtained when
the set of equations 3.2 and 3.3 are satisfied simultaneously.

$$v_k = \frac{1}{\sum_{k=1}^{n}(\mu_{ki})^m}\sum_{i=1}^{n}(\mu_{ki})^m x_i \qquad m > 1 \quad k = 1,2,\ldots,c \qquad (3.2)$$

$$\mu_{ki} = \frac{\left(\frac{1}{\|x_i - v_k\|^2}\right)^{1/m-1}}{\sum_{j=1}^{c}\left(\frac{1}{\|x_i - v_k\|^2}\right)^{1/m-1}}, \qquad k = 1,2,\ldots,c; i = 1,2,\ldots,n \qquad (3.3)$$

The fuzzy c-means algorithm solves these equations iteratively, to converge to a (perhaps local) optimum value of (3.1). It should be noted that the number of groups, c, is not estimated from the data, but must be specified before the algorithm is applied. However, we have found that a small range of values can be tried to choose the most suitable number of groups (see section 4). Also, it is necessary to specify the norm on X, $\|.\|$, by choosing a suitable symmetric, positive definite (pxp) matrix G. This matrix indicates the relative importance of each element, and correlations between them. Examples of G which could be used are the identity matrix, a diagonal matrix (with appropriate terms) and the covariance matrix, $Cov(x_i)$. The norm is then defined by

$$\|x\|_G = x'Gx$$

The algorithm may now be summarised as follows:

Step 1

Specify the number of clusters, c ($2 \le c \le n$),
the parameter m ($m > 1$),
the matrix G,
a small positive number, e, to measure the convergence.

Initialise the membership function values, M as $\mathrm{M}^{(0)}$,
the counter $l = 0$.

Step 2

Calculate the centres of the fuzzy clusters, $\{\upsilon_k^{(l)}\}, k = 1,2,\ldots,c$, using $\mathrm{M}^{(l)}$ and equation (3.2).

Step 3

Calculate the $\mathrm{M}^{(l+1)}$, using $\{\upsilon_k^{(l)}\}$ and equation (3.3) if $x_i \neq \upsilon_k^{(l)}$. Otherwise, set

$$\mu_{kj} = \begin{cases} 1 & \text{for } j = i \\ 0 & \text{for } j \neq i \end{cases}.$$

Step 4

Calculate $\Delta = \left\| \mathrm{M}^{(l+1)} - \mathrm{M}^{(l)} \right\|_G$.

If $\Delta > e$, then set $l = l+1$ and return to step 2.

This algorithm will converge to a local optimum, and care must be taken to check the results using different initial partitions to ensure that the result is consistent. There are no computational problems, even when there are a large number of elements.

The final problem is to consider the validity of the clustering which has resulted from the fuzzy c-means algorithm. Part of this involves choosing an appropriate value for c, and the criteria for this are discussed by Bezdek (1981). The illustrations in the next section provide some indications of the way in which this choice may be approached. Also, the results of the clustering have to be interpreted in relation to the aim of selecting groups for the policyholder age factor. This is covered in section 4.

4. Example

For this example we use a set of data based on more than 50,000 motor policies. For each policy, we have, for each of 2 types of claims (material damage, *md*, and bodily injury, *bi*), the number of claims, the total cost of the claims and the earned driver years (exposure). The data for the youngest ages (<25) and oldest ages (>82) were grouped in order to achieve exposures greater than 30. Note that the for this data set, the exposures were low at low ages, making this initial grouping advisable: a more comprehensive data set, or one with a different distribution over exposure, would remove the need for this.

Firstly, two quantities of interest are derived from the data set, the frequency and severity for *md* and *bi* claims. A summary of these values is given in Table 1. It is necessary at this stage to remove distortion effects due to the uneven mix of business by policyholder age. i.e. to calculate standardised frequencies and severities after having taken into account other factors such as car group, sex, etc. Standard techniques, such as generalised linear models (see Brockman and Wright, 1991), or an approach similar to that employed by Taylor (1989) can be used. This involves the use of well-documented estimation methods, keeping the ages separate. However, in this example, it can be assumed that this has already been done and that policyholder age is the only significant factor. This allows us to concentrate on the application of the fuzzy clustering techniques.

Considering the data in Table 1, it was found that the claim frequency seemed to give a reasonable indication of the variability of the risk by policyholder age, but that claim severity did not show any clear pattern and contained a high degree of variability. However, *bi* claims were, on average, more than 13 times as costly as *md* claims. For this reason, the fuzzy c-means algorithm will be applied to the claim frequencies and to adjusted claim frequencies, defined as claim frequencies multiplied by the average claim severity for each type of claim. Thus,

md adjusted frequency = *md* frequency x *md* severity and
bi adjusted frequency = *bi* frequency x *bi* severity

The claim frequencies and the adjusted claim frequencies for each type of claim are shown in Table 1. Also shown are the crude premiums, calculated as follows:

crude premium = *md* adjusted frequency + *bi* adjusted frequency.

Although both unadjusted and adjusted frequencies were considered, the adjusted frequencies were expected to be superior. Note that this adjustment could have been incorporated into the norm matrix, *G*.

Table 1: *age*	*frequency md*	*frequency bi*	*severity md*	*severity bi*	*adjusted frequency md*	*adjusted frequency bi*	*crude premium*	*exposure*
<25	0.30691	0.04384	515.90	5381.68	121.10	248.14	369.25	43.54
25	0.27846	0.05967	439.60	2242.76	109.88	337.70	447.58	31.99
26	0.13580	0.01598	302.74	9742.90	53.59	90.42	144.01	79.66
27	0.18732	0.01767	364.29	4286.98	73.91	100.01	173.93	360.11
28	0.20380	0.01002	395.48	5254.42	80.42	56.73	137.15	571.41
29	0.18907	0.01126	362.36	4598.03	74.61	63.75	138.35	790.95
30	0.18175	0.01486	434.33	6041.05	71.72	84.10	155.82	1113.44
31	0.14277	0.01142	406.46	7614.30	56.34	64.64	120.98	1671.45
32	0.15469	0.00729	331.00	5928.38	61.04	41.26	102.30	2007.57
33	0.12644	0.00651	300.71	5423.97	49.89	36.85	86.74	1857.12
34	0.12914	0.00861	416.71	6037.58	50.96	48.73	99.69	1921.73
35	0.14105	0.00641	369.25	6043.16	55.66	36.29	91.94	1885.85
36	0.12895	0.00794	414.31	4742.69	50.88	44.96	95.85	2082.56
37	0.14444	0.00698	365.89	5473.63	57.00	39.53	96.52	2004.59
38	0.12641	0.00967	423.00	5445.77	49.88	54.74	104.62	1907.93
39	0.12772	0.00872	458.40	4757.70	50.40	49.37	99.77	1823.65
40	0.12218	0.00732	356.89	3381.56	48.21	41.40	89.61	1739.68
41	0.11796	0.01222	422.91	3424.66	46.55	69.14	115.69	1666.95
42	0.11471	0.00828	401.68	6085.11	45.26	46.85	92.11	1614.38
43	0.11017	0.00646	416.18	4804.83	43.47	36.55	80.02	1675.11
44	0.11005	0.00758	385.92	5724.75	43.42	42.88	86.30	1596.01
45	0.11247	0.00287	345.49	4551.51	44.38	16.26	60.64	1550.33
46	0.11479	0.00776	360.96	3789.87	45.29	43.89	89.19	1640.99
47	0.11970	0.00874	385.04	7947.82	47.23	49.45	96.68	1456.71
48	0.11975	0.01023	411.26	3776.65	47.25	57.88	105.13	1493.30
49	0.12098	0.01052	343.41	4671.31	47.74	59.54	107.27	1209.86
50	0.11047	0.00440	337.79	8008.71	43.59	24.90	68.49	1301.91
51	0.14009	0.01302	386.38	4150.16	55.28	73.69	128.96	1221.93
52	0.12067	0.00546	396.67	7867.67	47.61	30.90	78.52	1165.49
53	0.12340	0.00507	391.90	2685.65	48.69	28.70	77.40	1129.33
54	0.12615	0.00788	455.80	6436.11	49.78	44.62	94.40	887.81
55	0.10273	0.00654	535.29	6999.40	40.54	37.03	77.57	972.51
56	0.09071	0.00542	366.74	5385.20	35.79	30.65	66.44	940.04
57	0.07957	0.00918	387.84	3859.75	31.40	51.96	83.36	415.89
58	0.11671	0.00824	382.74	2726.71	46.05	46.63	92.68	463.46
59	0.10629	0.00518	496.70	11538.43	41.94	29.34	71.28	490.95
60	0.07552	0.00252	370.90	15990.23	29.80	14.25	44.05	505.56
61	0.08699	0.00829	619.72	9716.70	34.33	46.89	81.22	460.85
62	0.08507	0.00597	503.03	3677.61	33.57	33.79	67.36	426.37
63	0.06762	0.00432	353.15	3397.81	26.68	24.43	51.11	442.31
64	0.06307	0.00293	696.58	27154.86	24.89	16.60	41.49	433.88
65	0.07488	0.00326	346.98	2545.45	29.55	18.42	47.97	390.95
66	0.08823	0.01307	514.26	6541.60	34.81	73.97	108.79	389.50
67	0.08404	0.00615	343.60	4642.38	33.16	34.80	67.96	310.47
68	0.08037	0.00423	304.66	1415.02	31.71	23.94	55.65	300.88
69	0.08252	0.00236	350.23	23768.18	32.56	13.34	45.91	269.91
70	0.06287	0.00286	310.02	7089.09	24.81	16.17	40.98	222.70
71	0.07759	0.00517	749.87	9937.03	30.62	29.27	59.89	246.06
72	0.07509	0.00901	401.62	3782.42	29.63	51.00	80.63	211.86
73	0.08597	0.00000	351.00	0.00	33.92	0.00	33.92	170.25
74	0.04838	0.00000	430.71	0.00	19.09	0.00	19.09	157.85
75	0.04505	0.00000	379.38	0.00	17.77	0.00	17.77	155.39
76	0.09957	0.01494	378.72	3991.06	39.29	84.53	123.82	127.82
77	0.06820	0.00000	662.06	0.00	26.91	0.00	26.91	74.65
78	0.11220	0.00863	276.11	636.36	44.27	48.85	93.12	73.73
79	0.04088	0.00000	172.09	0.00	16.13	0.00	16.13	46.70
80	0.05759	0.01920	857.47	6532.27	22.73	108.65	131.38	33.15
81	0.06099	0.02033	1782.07	636.36	24.07	115.06	139.13	31.30
82	0.04193	0.00000	193.53	0.00	16.55	0.00	16.55	30.35
83+	0.00854	0.00000	396.89	0.00	3.37	0.00	3.37	74.50

The algorithm outlined in section 3 was applied, with the number of clusters, c, as 5, 6,7 and 10. A more rigorous method would have been to use optimality measures (Bezdek, 1981) to determine a suitable value for c, but we found the ad hoc approach to be sufficient. After studying the results, $c = 6$ was chosen as being the most suitable value, and the conclusions for the age groupings are summarised below.

Table 2 contains the membership values for the unadjusted frequencies and Table 3 shows the same results after applying a 20% cut. Note that this helps to clarify the results and makes the interpretation easier. The centres of the 6 clusters were:

Clusters	1	2	3	4	5	6
md frequency	0.292859	0.189539	0.134440	0.114335	0.079628	0.045360
bi frequency	0.051603	0.013503	0.009412	0.007597	0.005616	0.002293
Total frequency	0.344461	0.203041	0.143852	0.121932	0.085243	0.047653

Each centre is a two-dimensional vector and the third row is the sum of the two components.

Table 4 shows the membership values for the adjusted frequencies, after applying a 20% cut. In this case the centres of the clusters and the crude premiums for each cluster were:

Cluster	1	2	3	4	5	6
md adjusted frequency	114.48	48.39	52.16	48.53	36.80	21.94
bi adjusted frequency	292.07	90.79	64.61	43.44	29.85	4.47
Crude premium	406.55	139.18	116.77	91.97	66.65	26.41

Table 2: *cluster*

age	*1*	*2*	*3*	*4*	*5*	*6*
<25	0.96053	0.01684	0.00800	0.00644	0.00466	0.00353
25	0.94411	0.02563	0.01106	0.00868	0.00606	0.00446
26	0.00155	0.01386	0.89201	0.07550	0.01229	0.00479
27	0.00179	0.98356	0.00766	0.00404	0.00187	0.00108
28	0.01999	0.89561	0.04013	0.02410	0.01251	0.00767
29	0.00042	0.99622	0.00174	0.00093	0.00043	0.00025
30	0.00433	0.94781	0.02615	0.01290	0.00564	0.00316
31	0.00264	0.02910	0.86827	0.07746	0.01586	0.00666
32	0.01149	0.19305	0.58377	0.14857	0.04292	0.02020
33	0.00157	0.01156	0.64341	0.31516	0.02124	0.00707
34	0.00087	0.00677	0.86603	0.11273	0.01009	0.00352
35	0.00188	0.01964	0.89489	0.06595	0.01250	0.00514
36	0.00095	0.00740	0.84832	0.12819	0.01124	0.00390
37	0.00365	0.04222	0.82768	0.09668	0.02086	0.00891
38	0.00147	0.01085	0.67236	0.28908	0.01968	0.00655
39	0.00121	0.00919	0.77329	0.19591	0.01521	0.00518
40	0.00136	0.00923	0.27280	0.68619	0.02329	0.00713
41	0.00092	0.00580	0.10639	0.86171	0.01964	0.00554
42	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
43	0.00050	0.00277	0.02945	0.94424	0.01886	0.00417
44	0.00049	0.00273	0.02903	0.94493	0.01869	0.00412
45	0.00068	0.00393	0.04530	0.92287	0.02192	0.00529
46	0.00000	0.00000	0.00000	1.00000	0.00000	0.00000
47	0.00081	0.00525	0.11804	0.85538	0.01592	0.00462
48	0.00095	0.00616	0.13883	0.83025	0.01844	0.00537
49	0.00126	0.00834	0.21512	0.74585	0.02265	0.00679
50	0.00066	0.00368	0.03891	0.92675	0.02450	0.00550
51	0.00164	0.01668	0.90733	0.05887	0.01099	0.00449
52	0.00111	0.00736	0.17229	0.79201	0.02100	0.00623
53	0.00167	0.01163	0.36768	0.58355	0.02698	0.00848
54	0.00152	0.01116	0.63637	0.32324	0.02082	0.00689
55	0.00245	0.01234	0.09233	0.68962	0.17499	0.02827
56	0.00211	0.00921	0.04697	0.16092	0.73696	0.04383
57	0.00026	0.00102	0.00411	0.01022	0.97420	0.01017
58	0.00018	0.00111	0.01870	0.97459	0.00427	0.00115
59	0.00157	0.00828	0.07154	0.82148	0.08155	0.01558
60	0.00050	0.00191	0.00711	0.01633	0.94664	0.02750
61	0.00120	0.00506	0.02367	0.07126	0.86868	0.03013
62	0.00062	0.00254	0.01138	0.03245	0.93550	0.01752
63	0.00196	0.00695	0.02314	0.04739	0.71255	0.20800
64	0.00245	0.00840	0.02634	0.05106	0.48074	0.43101
65	0.00053	0.00200	0.00740	0.01684	0.94278	0.03044
66	0.00226	0.00953	0.04551	0.13744	0.75520	0.05006
67	0.00041	0.00169	0.00740	0.02053	0.95747	0.01250
68	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000
69	0.00039	0.00158	0.00664	0.01754	0.96065	0.01320
70	0.00245	0.00839	0.02626	0.05077	0.47003	0.44210
71	0.00009	0.00034	0.00133	0.00319	0.99093	0.00413
72	0.00061	0.00229	0.00854	0.01950	0.93669	0.03237
73	0.00136	0.00565	0.02529	0.07151	0.85892	0.03726
74	0.00023	0.00070	0.00188	0.00320	0.01397	0.98003
75	0.00008	0.00025	0.00066	0.00110	0.00434	0.99357
76	0.00368	0.01760	0.11435	0.52436	0.29403	0.04599
77	0.00213	0.00760	0.02529	0.05179	0.69834	0.21485
78	0.00016	0.00092	0.01117	0.98135	0.00517	0.00123
79	0.00037	0.00111	0.00279	0.00453	0.01612	0.97507
80	0.00410	0.01324	0.03849	0.06887	0.34477	0.53053
81	0.00443	0.01464	0.04400	0.08067	0.43032	0.42594
82	0.00025	0.00076	0.00193	0.00315	0.01150	0.98241
83+	0.01064	0.02697	0.05574	0.07898	0.17474	0.65292

Table 3: *cluster*	*1*	*2*	*3*	*4*	*5*	*6*
age						
<25	1.00	0.00	0.00	0.00	0.00	0.00
25	1.00	0.00	0.00	0.00	0.00	0.00
26	0.00	0.00	1.00	0.00	0.00	0.00
27	0.00	1.00	0.00	0.00	0.00	0.00
28	0.00	1.00	0.00	0.00	0.00	0.00
29	0.00	1.00	0.00	0.00	0.00	0.00
30	0.00	1.00	0.00	0.00	0.00	0.00
31	0.00	0.00	1.00	0.00	0.00	0.00
32	0.00	0.00	1.00	0.00	0.00	0.00
33	0.00	0.00	0.67	0.33	0.00	0.00
34	0.00	0.00	1.00	0.00	0.00	0.00
35	0.00	0.00	1.00	0.00	0.00	0.00
36	0.00	0.00	1.00	0.00	0.00	0.00
37	0.00	0.00	1.00	0.00	0.00	0.00
38	0.00	0.00	0.70	0.30	0.00	0.00
39	0.00	0.00	1.00	0.00	0.00	0.00
40	0.00	0.00	0.28	0.72	0.00	0.00
41	0.00	0.00	0.00	1.00	0.00	0.00
42	0.00	0.00	0.00	1.00	0.00	0.00
43	0.00	0.00	0.00	1.00	0.00	0.00
44	0.00	0.00	0.00	1.00	0.00	0.00
45	0.00	0.00	0.00	1.00	0.00	0.00
46	0.00	0.00	0.00	1.00	0.00	0.00
47	0.00	0.00	0.00	1.00	0.00	0.00
48	0.00	0.00	0.00	1.00	0.00	0.00
49	0.00	0.00	0.22	0.78	0.00	0.00
50	0.00	0.00	0.00	1.00	0.00	0.00
51	0.00	0.00	1.00	0.00	0.00	0.00
52	0.00	0.00	0.00	1.00	0.00	0.00
53	0.00	0.00	0.39	0.61	0.00	0.00
54	0.00	0.00	0.66	0.34	0.00	0.00
55	0.00	0.00	0.00	1.00	0.00	0.00
56	0.00	0.00	0.00	0.00	1.00	0.00
57	0.00	0.00	0.00	0.00	1.00	0.00
58	0.00	0.00	0.00	1.00	0.00	0.00
59	0.00	0.00	0.00	1.00	0.00	0.00
60	0.00	0.00	0.00	0.00	1.00	0.00
61	0.00	0.00	0.00	0.00	1.00	0.00
62	0.00	0.00	0.00	0.00	1.00	0.00
63	0.00	0.00	0.00	0.00	0.77	0.23
64	0.00	0.00	0.00	0.00	0.53	0.47
65	0.00	0.00	0.00	0.00	1.00	0.00
66	0.00	0.00	0.00	0.00	1.00	0.00
67	0.00	0.00	0.00	0.00	1.00	0.00
68	0.00	0.00	0.00	0.00	1.00	0.00
69	0.00	0.00	0.00	0.00	1.00	0.00
70	0.00	0.00	0.00	0.00	0.52	0.48
71	0.00	0.00	0.00	0.00	1.00	0.00
72	0.00	0.00	0.00	0.00	1.00	0.00
73	0.00	0.00	0.00	0.00	1.00	0.00
74	0.00	0.00	0.00	0.00	0.00	1.00
75	0.00	0.00	0.00	0.00	0.00	1.00
76	0.00	0.00	0.00	0.64	0.36	0.00
77	0.00	0.00	0.00	0.00	0.76	0.24
78	0.00	0.00	0.00	1.00	0.00	0.00
79	0.00	0.00	0.00	0.00	0.00	1.00
80	0.00	0.00	0.00	0.00	0.39	0.61
81	0.00	0.00	0.00	0.00	0.50	0.50
82	0.00	0.00	0.00	0.00	0.00	1.00
83+	0.00	0.00	0.00	0.00	0.00	1.00

Table 4: *cluster* *age*	*1*	*2*	*3*	*4*	*5*	*6*
<25	1.00	0.00	0.00	0.00	0.00	0.00
25	1.00	0.00	0.00	0.00	0.00	0.00
26	0.00	1.00	0.00	0.00	0.00	0.00
27	0.00	0.69	0.31	0.00	0.00	0.00
28	0.00	0.00	0.56	0.44	0.00	0.00
29	0.00	0.00	0.66	0.34	0.00	0.00
30	0.00	0.51	0.49	0.00	0.00	0.00
31	0.00	0.00	1.00	0.00	0.00	0.00
32	0.00	0.00	0.00	1.00	0.00	0.00
33	0.00	0.00	0.00	0.76	0.24	0.00
34	0.00	0.00	0.00	1.00	0.00	0.00
35	0.00	0.00	0.00	0.76	0.24	0.00
36	0.00	0.00	0.00	1.00	0.00	0.00
37	0.00	0.00	0.00	1.00	0.00	0.00
38	0.00	0.00	0.38	0.62	0.00	0.00
39	0.00	0.00	0.00	1.00	0.00	0.00
40	0.00	0.00	0.00	1.00	0.00	0.00
41	0.00	0.00	1.00	0.00	0.00	0.00
42	0.00	0.00	0.00	1.00	0.00	0.00
43	0.00	0.00	0.00	0.44	0.56	0.00
44	0.00	0.00	0.00	1.00	0.00	0.00
45	0.00	0.00	0.00	0.00	0.71	0.29
46	0.00	0.00	0.00	1.00	0.00	0.00
47	0.00	0.00	0.00	1.00	0.00	0.00
48	0.00	0.00	0.61	0.39	0.00	0.00
49	0.00	0.00	0.73	0.27	0.00	0.00
50	0.00	0.00	0.00	0.00	1.00	0.00
51	0.00	0.00	1.00	0.00	0.00	0.00
52	0.00	0.00	0.00	0.39	0.61	0.00
53	0.00	0.00	0.00	0.37	0.63	0.00
54	0.00	0.00	0.00	1.00	0.00	0.00
55	0.00	0.00	0.00	0.28	0.72	0.00
56	0.00	0.00	0.00	0.00	1.00	0.00
57	0.00	0.00	0.23	0.44	0.33	0.00
58	0.00	0.00	0.00	1.00	0.00	0.00
59	0.00	0.00	0.00	0.00	1.00	0.00
60	0.00	0.00	0.00	0.00	0.32	0.68
61	0.00	0.00	0.00	0.54	0.46	0.00
62	0.00	0.00	0.00	0.00	1.00	0.00
63	0.00	0.00	0.00	0.00	0.73	0.27
64	0.00	0.00	0.00	0.00	0.30	0.70
65	0.00	0.00	0.00	0.00	0.54	0.46
66	0.00	0.38	0.62	0.00	0.00	0.00
67	0.00	0.00	0.00	0.00	1.00	0.00
68	0.00	0.00	0.00	0.00	1.00	0.00
69	0.00	0.00	0.00	0.00	0.37	0.63
70	0.00	0.00	0.00	0.00	0.28	0.72
71	0.00	0.00	0.00	0.00	1.00	0.00
72	0.00	0.00	0.00	0.52	0.48	0.00
73	0.00	0.00	0.00	0.00	0.00	1.00
74	0.00	0.00	0.00	0.00	0.00	1.00
75	0.00	0.00	0.00	0.00	0.00	1.00
76	0.00	0.73	0.27	0.00	0.00	0.00
77	0.00	0.00	0.00	0.00	0.00	1.00
78	0.00	0.00	0.00	1.00	0.00	0.00
79	0.00	0.00	0.00	0.00	0.00	1.00
80	0.00	1.00	0.00	0.00	0.00	0.00
81	0.00	1.00	0.00	0.00	0.00	0.00
82	0.00	0.00	0.00	0.00	0.00	1.00
83+	0.00	0.00	0.00	0.00	0.00	1.00

Note that the centres of the clusters above have been calculated using the data for each age which might be in each cluster (i.e. for which the membership value is > 0). It is required by the underwriting procedures that the fuzzy results are converted into crisp age groups. This requires us to decide which group each individual age should be assigned to, using the results above. Deciding on the borders and sizes of each group is not straightforward, and it is necessary to take into account the following considerations. We would expect and require that risk should progress smoothly with age, although some variability will always be present. Thus, the age groups should contain only adjacent ages, and we will interpret the results of the fuzzy c-means algorithm to this effect. Considering the results for the adjusted frequencies in Table 4, which are considered to be more reliable, we define the age groupings as shown below.

Group	1	2	3	4	5	6	7
Risk Cluster	1	2	3	4	3	5	7
Ages	(<25,25)	(26,27)	(28,31)	(32,47)	(48,51)	(52,68)	(69,>69)

Clearly, group 5 is questionable since it indicates a higher risk compared to group 4 (as indicated by the risk cluster), and makes an unexpected progression in the risk rating. It is possible that this effect is real, but otherwise groups 4 and 5 could be amalgamated.

In order to assess whether adjacent groups should be amalgamated, and in order to assess the implications for the premium, we can calculate the following risk measure:

$$R_i = \frac{1}{\|i\|} \sum_{j=\text{ages in i}} \sum_{k=1}^{c} \mu_{kj} \|\upsilon_k\| \qquad \text{for each age group, } i.$$

The value of this measure for the groups above are:

Group, i	1	2	3	4	5	6	7
R_i	406.29	135.65	114.79	90.15	100.15	71.78	60.92

These values can be used to measure the relative risk of each group. For example, it can be seen that the highest risk group (group 1) has a risk measure which is nearly 7 times that of the lowest risk group (group 7). These values are presented graphically in Figure 1, which also shows the crude risk premiums for comparison purposes. For greater clarity, Figure 2 reproduces the results shown in Figure 1, omitting cluster 1. Although an analysis of the residuals is not appropriate, these figures can be used to identify any strange results which can be investigated further.

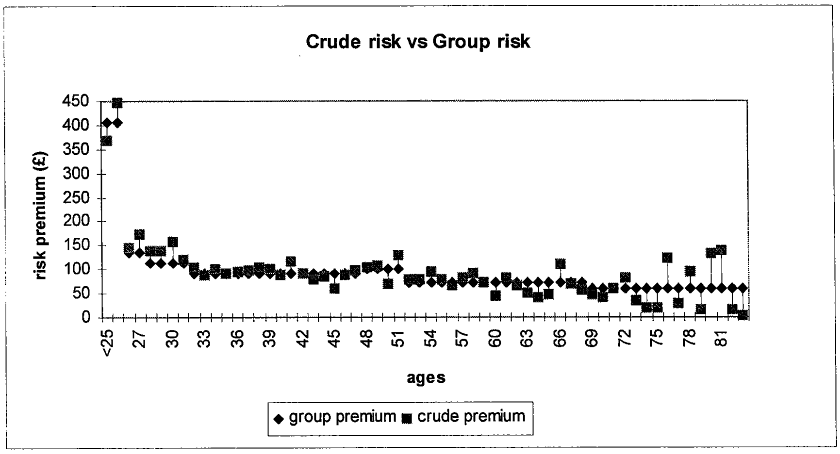

Figure 1. Comparison of the crude risk premium and the risk premium based on the risk groups.

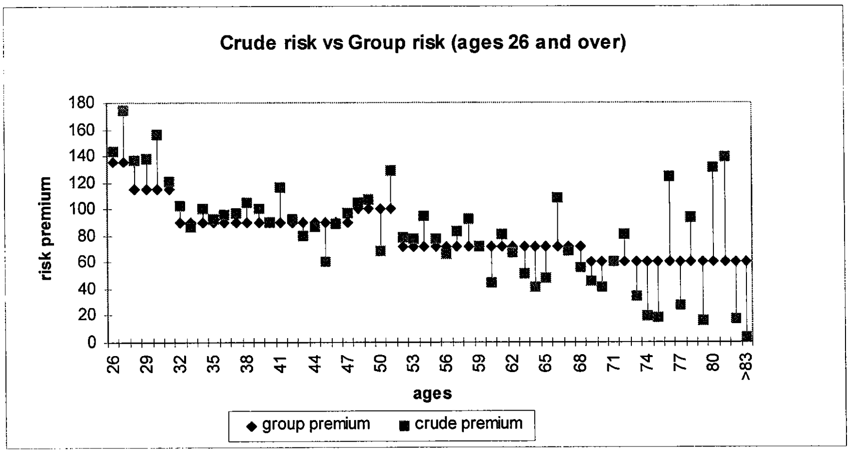

Figure 2: Comparison of the crude risk premium and the risk premium based on the risk groups (risk groups 2 to 6).

5. Conclusion

This paper has shown how the fuzzy c-means algorithm can be used to investigate age groupings in general insurance. The fuzzy approach is particularly suited to this problem, but there are some other methods which could also be used.

The obvious candidates are parametric and non-parametric smoothing within the framework of generalised linear models. These involve using a generalised linear model to fit the data, but treating the policyholder age as a continuous variable not as a factor. Thus, it would be possible to use, for example, a polynomial function of age to model the effect of policyholder age on the risk. It would also be possible to apply non-parametric smoothing methods, such as cubic smoothing splines, if a parametric model was not suitable. This makes the problem very similar to that of graduating life tables, and the reader is referred to Renshaw (1991) and Verrall (1996) for further discussion of these methods.

Closer in spirit to the approach taken in this paper would be the use of (crisp) clustering methods such as the minimum variance method (see van Eeghen *et al*, 1983), or the method proposed by Loimaranta *et al* (1980). The first of these two methods uses an algorithm which aims to break the data set up into a number of clusters, so that the within-cluster variance is small and the between-cluster variance is large. The second method derives the posterior probability that each data point belongs to each cluster, using a Bayesian approach. However, we believe that the flexibility of the fuzzy approach makes it more suitable for grouping policyholder age than any of the above methods.

The problem of grouping by policyholder age has not been considered previously in the actuarial literature, and we believe that it forms an important part of the underwriting process. This paper has shown how the fuzzy c-means algorithm can be used to assess the groups used in practice, and we recommend that an investigation of this type should be part of any risk rating exercise for a general insurance portfolio.

This paper has emphasised the application of the fuzzy c-means algorithm to grouping by policyholder age, but could also be applied to other explanatory variables and in other types of insurance. For example, the classification of vehicles into vehicle rating groups, the grouping of car engine sizes and the classification of excess mortality risk in life insurance according to blood pressure are all problems to which the approach described in this paper could be used.

References

Bezdek, J.C., 1981, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, London.

Brockman, M.J. and T.S. Wright, 1992, *Statistical motor rating: making effective use of your data*, J.I.A., **119**, 457-543.

Cummins, J.D. and R.A. Derrig, 1993, *Fuzzy trends in property - liability insurance claim costs*, Journal of Risk and Insurance, **60**, 429-465.

Cummins, J.D. and R.A. Derrig, 1997, *Fuzzy financial pricing of property liability insurance*, North-American Actuarial Journal, **1**, 21-40.

Derrig, R.A. and K.M. Ostaszewski, 1995, *Fuzzy techniques of pattern recognition in risk and claim classification*, Journal of Risk and Insurance, **62**, 447-482.

Derrig, R.A. and K.M. Ostaszewski, 1995, *The fuzzy problem of hedging the tax liability of a property - liability insurance company*, 5th AFIR International Colloquium, **1**, 17 - 42.

DeWit, G.W., 1982, *Underwriting and uncertainty*, Insurance: Mathematics and Economics, **1**, 277-285.

van Eeghen, J., E.K. Greup and J.A. Nijssen, 1983, *Rate making*, Nationale-Nederlanden N.V., The Netherlands.

Haberman, S. and A.E. Renshaw, 1996, *Generalized linear models and actuarial science*, J.R.S.S (Series D), **45**, 407-436.

Horgby, P.J., R. Lohse, N.A. Sittaro, 1997, *Fuzzy underwriting: an application of fuzzy logic to medical underwriting*, Journal of Actuarial Practice, **5**, 79-104.

Lemaire, J., 1990, *Fuzzy insurance*, ASTIN Bulletin, **20**, 33-55.

Loimaranta, K., J. Jacobsson and H. Lonka, 1980, *On the use of mixture models in clustering multivariate frequency data*, TICA, **21**, 147- 161.

Ostaszewski, K.M., 1993, *An investigation into possible applications of fuzzy set methods in actuarial science*, Society of Actuaries.

Renshaw, A.E., 1991, *Actuarial graduation practice and generalised linear and non-linear models*, J.I.A., **118**, 295-312.

Taylor, G.C., 1989, *Use of spline functions for premium rating by geographic area*, ASTIN Bulletin, **19**, 91-122.

Verrall, R.J., 1996, *A unified framework for graduation*, Actuarial Research Report No. 91, City University, London.

Zadeh, L. A., 1965, *Fuzzy sets*, Information and Control, **8**, 338-353.

Zimmerman, H.J., 1991, *Fuzzy set theory and its applications*, Kluwer Academic Publishers, Boston/Dordrecht/London.

Yakoubov, Y.H. and S. Haberman, 1998, *Review of actuarial applications of fuzzy set theory,* Actuarial Research Report, City University, London, forthcoming.

CITY University

ACTUARIAL RESEARCH CENTRE

DEPARTMENT OF ACTUARIAL SCIENCE AND STATISTICS

Actuarial Research Papers

1. Haberman S. Long Term Prognosis after Cerebral Infarction. 78 pages, 1984 (out of print).
ISBN 1 874 770 01 8

2. Haberman S. Long Term Prognosis after Subarachnoid Haemorrhage. 43 pages, 1984 (out of print).
ISBN 1 874 770 02 6

3. Haberman S. Age Variation in Incidence of Cerebrovascular Disease. 36 pages, 1985 (out of print).
ISBN 1 874 770 03 4

4. Kaye G.D. Taxation of Pension Funds. 11 pages, 1985.
ISBN 1 874 770 04 2

[Also: Haberman S. and Bloomfield D.S.F. Social Class Differences in Mortality Around 1981. SSRU Working Paper No. 55. 41 pages, 1987.]

5. Kaye G.D. Life Offices' Expenses Regression Analyses I, 98 pages, 1988. (Limited circulation).
ISBN 1 874 770 05 0

6. Kaye G.D. Life Assurance Taxation - Comments in Response to the Board of the Inland Revenue's Consultative Document Concerning the Taxation of Life Insurance. 11 pages. 1988.
ISBN 1 874 770 06 9

7. Verrall R.J. Bayesian Linear Models and the Claims Run-Off Triangle. 32 pages. 1988.
ISBN 1 874 770 07 7

8. Kaye G.D. Life Offices' Expenses Regression Analyses II. 12 pages. 1989. (Limited circulation).
ISBN 1 874 770 08 5

9. Bloomfield D.S.F. and Haberman S. The Assessment of Damages and the Measurement of Work time Lost to Sickness, Unemployment and Stoppages. 100 pages. 1989.
ISBN 1 874 770 09 3

10. Adams A.T. Pricing Property Investment Company Shares. 14 pages. 1989.
ISBN 1 874 770 10 7

11. Booth P.M. History of Exchange Controls in the EC with Particular Reference to the United Kingdom. 39 pages. 1989.
ISBN 1 874 770 11 5

12. Verrall R.J. Chain Ladder and Maximum Likelihood. 12 pages. 1989.
ISBN 1 874 770 12 3

13. Verrall R.J. On the Unbiased Estimation of Reserves from Loglinear Models. 13 pages, 1989.
ISBN 1 874 770 13 1

14. Adams A.T. An Analysis of Investment Trust Shares, 1990. 31 pages.
ISBN 1 874 770 14 X

15. Haberman S. Demographic Trends Towards 2000. 16 pages. 1989.
ISBN 1 874 770 15 8

16. Haberman S. British Life Insurance: The Role of the Actuary and 1992. 30 pages. 1990.
ISBN 1 874 770 16 6

17. Verrall R.J. Bootstrapping, Saddlepoint Approximations and the Probability of Ruin. 10 pages. July 1990.
ISBN 1 874 770 17 4

18. Booth P.M. Theoretical Aspects of the Taxation of Real Returns from Short Term Deposits. 26 pages. June 1990.
ISBN 1 874 770 18 2

19. Booth P.M. Practical Aspects of the Taxation of Real Interest. 38 pages, 1990.
ISBN 1 874 770 19 0

20. Puzey A.S. Actuarial Funding in Unit-linked Life Assurance - Derivation of the Actuarial Funding Factors. 19 pages. July 1990.
ISBN 1 874 770 20 4

21. Kaye G.D. Economies of Scale in the United Kingdom Insurance Industry. A report for the Association of British Insurers. Limited Circulation. October 1990.
ISBN 1 874 770 21 2

22. Adams A.T. & Venmore-Rowland P. Direct or Indirect Property Investment. 71 pages. September 1990.
ISBN 1 874 770 22 0

23. England P.D. Statistical Modelling of Excess Mortality. 124 pages. September 1990.
ISBN 1 874 770 23 9

24. Verrall R.J. & Li Z. Negative Incremental Claims: Chain Ladder and Linear Models. November 1990.
ISBN 1 874 770 24 7

25. Haberman S. State Pension Provision and Social Security in the United Kingdom. Presented at a Conference in Lisbon on "Social Security and Private Provision; What are the Prospects?". 40 pages. October 1990.
ISBN 1 874 770 25 5

26. Renshaw A.E. Graduation by Generalized Linear Modelling Techniques. 42 pages. February 1991.
ISBN 1 874 770 26 3

27. England P.D. Statistical Modelling of Excess Mortality No. 2. 170 pages. June 1991.
ISBN 1 874 770 27 1

28. Adams, A.T. The Market Pricing of Composite Insurance Shares. 37 pages. July 1991.
ISBN 1 874 770 28 X

29. Haberman, S. The Effect of Autoregressive Rates of Return on the Variability of Pension Contributions and Fund Levels. 58 pages. September 1991.
ISBN 1 874 770 29 8

30. Puzey, A.S. An Investigation into the Effect of Averaging Age Nearest Birthday at the Beginning of a Rate Year to Exact Age. October 1991. 33 pages.
ISBN 1 874 770 30 1

31. Chadburn, R.G. Bias in Select Mortality Investigations where Data are Subdivided by Age at Death. November 1991. 69 pages.
ISBN 1 874 770 31 X

32. England P.D. and Verrall R.J. Dynamic Estimation for Models of Excess Mortality. January 1992. 19 pages.
ISBN 1 874 770 32 8

33. Verrall R.J. A State Space Formulation of Whittaker-Henderson Graduation, with Extensions. January 1992. 13 pages.
ISBN 1 874 770 33 6

34. Verrall R.J. Graduation by Dynamic Regression Methods. January 1992. 37 pages.
ISBN 1 874 770 34 4

35. Gerrard R.G. and Haberman S. Stability of Pension Systems when Gains/Losses are Amortized and Rates of Return are Autoregressive. March 1992. 12 pages.
ISBN 1 874 770 35 2

36. Haberman S. HIV, AIDS and the Approximate Calculation of Life Insurance Functions, Annuities and Net Premiums. April 1992. 28 pages.
ISBN 1 874 770 36 0

37. Cooper D.R. Savings and Loans: An Historical Perspective. May 1992. 29 pages.
ISBN 1 874 770 37 9

38. Verrall R.J. Dynamic Bayes Estimation for AIDS Delay Tables. May 1992. 16 pages.
ISBN 1 874 770 38 7

39. Kaye G.D. Data Sources Relating to Life Assurance Expenses. May 1992. 39 pages. Presented to the Staple Inn Actuarial Society and Royal Statistical Society - February 1992.
ISBN 1 874 770 39 5

40. Renshaw A.E., and Haberman S. On the Graduation Associated with a Multiple State Model for Permanent Health Insurance. May 1992.
ISBN 1 874 770 40 9

41. England P.D. Statistical Modelling of Excess Mortality Number 3. June 1992. 163 pages.
ISBN 1 874 770 41 7

42. Bloomfield D.S.F. and Haberman S. Male Social Class Mortality Differences Around 1981: An Extension to Include Childhood Ages. 21 pages. June 1992.
ISBN 1 874 770 42 5

43. Berg M.P and Haberman S. Trend Analysis and Prediction Procedures for Time Nonhomogeneous Claim Processes. 33 pages. June 1992.
ISBN 1 874 770 43 3

44. Booth P.M. The Single Market for Insurance, Free Capital Movements and attempts to Fix Exchange Rates. October 1992. 28 pages.
ISBN 1 874 770 44 1

45. Verrall R.J. Chain Ladder with Varying Run-off Evolutions. February 1993. 15 pages.
ISBN 1 874 770 45 X

46. Gavin J., Haberman S. and Verrall R.J. Moving Weighted Average Graduation using Kernel Estimation. November 1992. 14 pages.
ISBN 1 874 770 46 8

47. Gavin J., Haberman S. and Verrall R.J. On the Choice of Bandwidth for Kernel Graduation. November 1992. 21 pages.
ISBN 1 874 770 47 6

48. S. Haberman. Pension Funding with Time Delays and the Optimal Spread Period. May 1993. 13 pages.
ISBN 1 874 770 48 4

49. S. Haberman. Stochastic Investment Returns and the Present Value of Future Contributions in a Defined Benefit Pension Scheme. May 1993. 22 pages.
ISBN 1 874 770 49 2

50. A. Zimbidis and S. Haberman. Delay, Feedback and Variability of Pension Contributions and Fund Levels. May 1993. 19 pages.
ISBN 1 874 770 50 6

51. S. Haberman. Pension Funding: The Effect of Changing The Frequency Valuations. June 1993. 19 pages.
ISBN 1 874 770 51 4

52. S Haberman. HIV, AIDS Markov Chains and PHI. June 1993. 15 pages.
ISBN 1 874 770 52 2

53. S Haberman. A Consideration of Pension Credit and Termination Insurance. June 1993. 22 pages.
ISBN 1 874 770 53 0

54. M Z Khorasanee. Survey of Actuarial Practice in the Funding of UK Defined Benefit Pension Schemes. July 1993. 19 pages.
ISBN 1 874 770 54 9

55. P M Booth, R G Chadburn and A S K Ong. A Utility Maximisation Approach to Asset Allocation. September 1993. 40 pages.
ISBN 1 874 770 55 7

56. R G Chadburn. Bias in Select Mortality Investigations. August 1993. 62 pages.
ISBN 1 874 770 56 5

57. M Z Khorasanee. A Comparison of Pension Funding Strategies by means of Simulations for a Model Scheme. August 1993. 43 pages.
ISBN 1 874 770 57 3

58. A E Renshaw, P Hatzopolous and S Haberman. Recent Mortality Trends in Male Assured Lives. June 1993. 23 pages.
ISBN 1 874 770 58 1

59. E Pitacco. Disability Risk Models: Towards A Unifying Approach. September 1993. 33 pages.
ISBN 1 874 770 59 X

60. M Boskov and R J Verrall. Premium Rating by Geographic Area Using Spatial Models. September 1993. 14 pages.
ISBN 1 874 770 60 3

61. R G Chadburn. Managing Mutual Life Offices: The Use of an Expense Ratio in New Business Decision Making and Expense Control. October 1993. 21 pages.
ISBN 1 874 770 61 1

62. Haberman S. Pension Funding Modelling and Stochastic Investment Returns. 56 pages. March 1994.
ISBN 1 874 770 62 X

63. Renshaw A E and Verrall R J. The Stochastic Model Underlying the Chain-Ladder Technique. 25 pages. April 1994.
ISBN 1 874 770 63 8

64. Haberman S and Sung J-H. Dynamic Approaches to Pension Funding. 22 pages. April 1994.
ISBN 1 874 770 64 6

65. Renshaw A.E. On the Second Moment Properties and the Implementation of Certain GLIM Based Stochastic Claims Reserving Models. 36 pages. September 1994.
ISBN 1 874 770 65 4

66. Booth P.M., J.N. Allan, and J.W. Jang. An Evaluation of the UK Life Insurance Mismatching Test. September 1994.
ISBN 1 874 770 66 2

67. Booth P.M. and Stroinski K. Insurance and Investment Markets in Poland. September 1994. 35 pages.
ISBN 1 874 770 67 0

68. Ong A. A Stochastic Model for Treasury-Bills: An Extension to Wilkie's Model. September 1994. 12 pages.
ISBN 1 874 770 68 9

69. Bloomfield D.S.F. Moving on from Undergraduate Exams to Professional Exams: Actuaries. November 1994. 22 pages.
ISBN 1 874 770 69 7

70. Huber P. A Review of Wilkie's Stochastic Investment Model. January 1995. 22 pages.
ISBN 1 874 770 70 0

71. Renshaw A.E. On the Graduation of `Amounts'. January 1995. 24 pages.
ISBN 1 874 770 71 9

72. Renshaw A.E. Claims Reserving by Joint Modelling. December 1994. 26 pages.
ISBN 1 874 770 72 7

73. Renshaw A.E. Graduation and Generalised Linear Models: An Overview. February 1995. 40 pages.
ISBN 1 874 770 73 5

74. Khorasanee M.Z. Simulation of Investment Returns for a Money Purchase Fund. June 1995. 20 pages.
ISBN 1 874 770 74 3

75. Owadally M.I. and Haberman S. Finite-time Pension Fund Dynamics with Random Rates of Return. June 1995. 28 pages.
ISBN 1 874 770 75 1

76. Owadally M.I. and Haberman S. Stochastic Investment Modelling and Optimal Funding Strategies. June 1995. 25 pages.
ISBN 1 874 770 76 X

77. Khorasanee M.Z Applying the Defined Benefit Principle to a Defined Contribution Scheme. August 1995. 30 pages.
ISBN 1 874 770 77 8

78. Sebastiani P. and Settimi R. Experimental Design for Non-Linear Problems. September 1995. 13 pages.
ISBN 1 874 770 78 6

79. Verrall R.J. Whittaker Graduation and Parametric State Space Models. November 1995. 23 pages.
ISBN 1 874 770 79 4

80. Verrall R.J. Claims Reserving and Generalised Additive Models. November 1995. 17 pages.
ISBN 1 874 770 80 8

81. Nelder J.A. and Verrall R.J. Credibility Theory and Generalized Linear Models. November 1995. 15 pages.
ISBN 1 874 770 81 6

82. Renshaw A.E., Haberman S. and Hatzopoulos P. On The Duality of Assumptions Underpinning The Construction of Life Tables. December 1995. 17 Pages.
ISBN 1 874 770 82 4

83. Chadburn R.G. Use of a Parametric Risk Measure in Assessing Risk Based Capital and Insolvency Constraints for With Profits Life Insurance. March 1996. 17 Pages.
ISBN 1 874 770 84 0

84. Haberman S. Landmarks in the History of Actuarial Science (up to 1919). March 1996. 62 Pages.
ISBN 1 874 770 85 9

85. Renshaw A.E. and Haberman S. Dual Modelling and Select Mortality. March 1996. 30 Pages.
ISBN 1 874 770 88 3

86. Booth P.M. Long-Term Care for the Elderly: A Review of Policy Options. April 1996. 45 Pages.
ISBN 1 874 770 89 1

87. Huber P.P. A Note on the Jump-Equilibrium Model. April 1996. 17 Pages.
ISBN 1 874 770 90 5

88. Haberman S and Wong L.Y.P. Moving Average Rates of Return and the Variability of Pension Contributions and Fund Levels for a Defined Benefit Pension Scheme. May 1996. 51 Pages.
ISBN 1 874 770 91 3

89. Cooper D.R. Providing Pensions for Employees with Varied Working Lives. June 1996. 25 Pages.
ISBN 1 874 770 93 X

90. Khorasanee M.Z. Annuity Choices for Pensioners. August 1996. 25 Pages.
ISBN 1 874 770 94 8

91. Verrall R.J. A Unified Framework for Graduation. November 1996. 25 Pages.
ISBN 1 874 770 99 9

92. Haberman S. and Renshaw A.E. A Different Perspective on UK Assured Lives Select Mortality. November 1996. 61 Pages.
ISBN 1 874 770 00 X

93. Booth P.M. The Analysis of Actuarial Investment Risk. March 1997. 43 Pages.
ISBN 1 901615 03 0

94. Booth P.M., Chadburn R.G. and Ong A.S.K. Utility-Maximisation and the Control of Solvency for Life Insurance Funds. April 1997. 39 Pages.
ISBN 1 901615 04 9

95. Chadburn R.G. The Use of Capital, Bonus Policy and Investment Policy in the Control of Solvency for With-Profits Life Insurance Companies in the UK. April 1997. 29 Pages.
ISBN 1 901615 05 7

96. Renshaw A.E. and Haberman S. A Simple Graphical Method for the Comparison of Two Mortality Experiences. April 1997. 32 Pages.
ISBN 1 901615 06 5

97. Wong C.F.W. and Haberman S. A Short Note on Arma (1, 1) Investment Rates of Return and Pension Funding. April 1997. 14 Pages.
ISBN 1 901615 07 3

98. Puzey A S. A General Theory of Mortality Rate Estimators. June 1997. 26 Pages.
ISBN 1 901615 08 1

99. Puzey A S. On the Bias of the Conventional Actuarial Estimator of q_x. June 1997. 14 Pages.
ISBN 1 901615 09 X

100. Walsh D. and Booth P.M. Actuarial Techniques in Pricing for Risk in Bank Lending. June 1997. 55 Pages.
ISBN 1 901615 12 X

101. Haberman S. and Walsh D. Analysis of Trends in PHI Claim Inception Data. July 1997. 51 Pages.
ISBN 1 901615 16 2

102. Haberman S. and Smith D. Stochastic Investment Modelling and Pension Funding: A Simulation Based Analysis. November 1997. 91 Pages.
ISBN 1 901615 19 7

103. Rickayzen B.D. A Sensitivity Analysis of the Parameters used in a PHI Multiple State Model. December 1997. 18 Pages.
ISBN 1 901615 20 0

104. Verrall R.J. and Yakoubov Y.H. A Fuzzy Approach to Grouping by Policyholder Age in General Insurance. January 1998. 18 Pages.
ISBN 1 901615 22 7

Statistical Research Papers

1. Sebastiani P. Some Results on the Derivatives of Matrix Functions. December 1995. 17 Pages.
ISBN 1 874 770 83 2

2. Dawid A.P. and Sebastiani P. Coherent Criteria for Optimal Experimental Design. March 1996. 35 Pages.
ISBN 1 874 770 86 7

3. Sebastiani P. and Wynn H.P. Maximum Entropy Sampling and Optimal Bayesian Experimental Design. March 1996. 22 Pages.
ISBN 1 874 770 87 5

4. Sebastiani P. and Settimi R. A Note on D-optimal Designs for a Logistic Regression Model. May 1996. 12 Pages.
ISBN 1 874 770 92 1

5. Sebastiani P. and Settimi R. First-order Optimal Designs for Non Linear Models. August 1996. 28 Pages.
ISBN 1 874 770 95 6

6. Newby M. A Business Process Approach to Maintenance: Measurement, Decision and Control. September 1996. 12 Pages.
ISBN 1 874 770 96 4

7. Newby M. Moments and Generating Functions for the Absorption Distribution and its Negative Binomial Analogue. September 1996. 16 Pages.
ISBN 1 874 770 97 2

8. Cowell R.G. Mixture Reduction via Predictive Scores. November 1996. 17 Pages.
ISBN 1 874 770 98 0

9. Sebastiani P. and Ramoni M. Robust Parameter Learning in Bayesian Networks with Missing Data. March 1997. 9 Pages.
ISBN 1 901615 00 6

10. Newby M.J. and Coolen F.P.A. Guidelines for Corrective Replacement Based on Low Stochastic Structure Assumptions. March 1997. 9 Pages.
ISBN 1 901615 01 4.

11. Newby M.J. Approximations for the Absorption Distribution and its Negative Binomial Analogue. March 1997. 6 Pages.
ISBN 1 901615 02 2

12. Ramoni M. and Sebastiani P. The Use of Exogenous Knowledge to Learn Bayesian Networks from Incomplete Databases. June 1997. 11 Pages.
ISBN 1 901615 10 3

13. Ramoni M. and Sebastiani P. Learning Bayesian Networks from Incomplete Databases. June 1997. 14 Pages.
ISBN 1 901615 11 1

14. Sebastiani P. and Wynn H.P. Risk Based Optimal Designs. June 1997. 10 Pages.
ISBN 1 901615 13 8

15. Cowell R. Sampling without Replacement in Junction Trees. June 1997. 10 Pages.
ISBN 1 901615 14 6

16. Dagg R.A. and Newby M.J. Optimal Overhaul Intervals with Imperfect Inspection and Repair. July 1997. 11 Pages.
ISBN 1 901615 15 4

17. Sebastiani P. and Wynn H.P. Bayesian Experimental Design and Shannon Information. October 1997. 11 Pages.
ISBN 1 901615 17 0

18. Wolstenholme L.C. A Characterisation of Phase Type Distributions. November 1997. 11 Pages.
ISBN 1 901615 18 9

19. Wolstenholme L.C. A Comparison of Models for Probability of Detection (POD) Curves. December 1997. 23 Pages.
ISBN 1 901615 21 9

ISBN 1 901615 22 7

Department of Actuarial Science and Statistics

Actuarial Research Club

The support of the corporate members

Commercial Union
Coopers & Lybrand
Government Actuary's Department
Guardian Insurance
Hymans Robertson
KPMG
Munich Reinsurance
Swiss Reinsurance

is gratefully acknowledged.